The Annals of Statistics
1976, Vol. 4, No. 2, 384-395

EXTENSION OF THE GAUSS-MARKOV THEOREM TO INCLUDE THE ESTIMATION OF RANDOM EFFECTS

BY DAVID HARVILLE

Aerospace Research Laboratories, Wright-Patterson AFB

The general mixed linear model can be written $y = X\alpha + Zb$, where α is a vector of fixed effects and b is a vector of random variables. Assume that $E(b) = 0$ and that $\text{Var}(b) = \sigma^2 D$ with D known. Consider the estimation of $\lambda_1'\alpha + \lambda_2'\beta$, where $\lambda_1'\alpha$ is estimable and β is the realized, though unobservable, value of b. Among linear estimators $c + r'y$ having $E(c + r'y) \equiv E(\lambda_1'\alpha + \lambda_2'b)$, mean squared error $E(c + r'y - \lambda_1'\alpha - \lambda_2'b)^2$ is minimized by $\lambda_1'\hat{\alpha} + \lambda_2'\hat{\beta}$, where $\hat{\beta} = DZ'V^{\#}(y - X\hat{\alpha})$, $\hat{\alpha} = (X'V^{\#}X)^{-}X'V^{\#}y$, and $V^{\#}$ is any generalized inverse of $V = ZDZ'$ belonging to the Zyskind-Martin class. It is shown that $\hat{\alpha}$ and $\hat{\beta}$ can be computed from the solution to any of a certain class of linear systems, and that doing so facilitates the exploitation, for computational purposes, of the kind of structure associated with ANOVA models. These results extend the Gauss-Markov theorem. The results can also be applied in a certain Bayesian setting.

1. Introduction. Most, if not all, linear statistical models that may be applied to a set of data are included in the formulation

$$y = X\alpha + Zb, \tag{1.1}$$

where y is a $n \times 1$ vector of random variables whose observed values comprise the data points, X and Z are matrices of known 'regressors' with dimensions $n \times p$ and $n \times q$ respectively, α is a $p \times 1$ vector of fixed unknown and unobservable parameters, and b is a $q \times 1$ vector of unobservable random effects or errors. It is assumed that $E(b) = 0$ and that $\text{Var}(b) = \sigma^2 D$ where σ^2 is a strictly positive parameter that is generally unknown and where the elements of the possibly singular matrix D are known functions of some vector θ of parameters. Define V by $V = ZDZ'$, so that $\text{Var}(y) = \sigma^2 V$. The above formulation includes, but is not limited to, the fixed, mixed, and random models associated with the analysis of variance.

In conjunction with the model (1.1), considerable attention has been given in the past to the problem of making inferences from the data about the parameter vectors α and/or θ. Much less emphasis has been devoted to the problem of making inferences about the *realized* or *sample value* β of the random vector b or, more generally, to the problem of making inferences about linear combinations of the elements of α and β. Yet, as observed by Searle (1974), it is the case in many applications that β enters nontrivially in the linear combinations of principal interest. For example, in animal breeding applications, linear

Received January 1974; revised August 1975.
AMS 1970 *subject classifications*. Primary 62J05, 62F10; Secondary 62F15, 15A09, 15A24.
Key words and phrases. Gauss-Markov theorem, linear models, mixed analysis-of-variance models, best linear unbiased estimation, singular covariance matrices, Bayesian inference.

combinations of the elements of α and b correspond to the breeding values of individual animals, and the primary objective in analyzing the data may be to evaluate these same individuals as candidates for some future breeding program.

In what follows, some results are presented on the estimation of linear combinations of the elements of α and β *for the case where θ and thus D are known.* These results are of interest even though, in practice, θ is often unknown. A good estimate of θ may be available or obtainable, in which case we can proceed as though D were in fact known. Moreover, knowing what to do when θ is known may provide some insight into what to do when it is not. The traditional advice on estimating estimable linear functions of α when θ is unknown has been to first confront the generally difficult problem of estimating θ, perhaps by using analysis-of-variance techniques if they are applicable, and to then proceed as though that estimate were the true θ-value. A similar approach can be adopted for the more general problem of estimating linear combinations of the elements of α and β. Also, the results derived on the basis of the model (1.1) for the case of known D are found to have an interesting interpretation in a Bayesian setting in which some elements of b are regarded as unknown parameters.

2. Best linear unbiased estimation of linear combinations of fixed and random effects. An estimator $t(y)$ of $\lambda_1'\alpha + \lambda_2'\beta$, where λ_1 is $p \times 1$ and λ_2 is $q \times 1$, will be called *unbiased* if $E[t(y)] \equiv E(\lambda_1'\alpha + \lambda_2'b) \equiv \lambda_1'\alpha$, and will be labelled *linear* if $t(y) \equiv c + r'y$ for some constant c and some $n \times 1$ vector r of constants. The quantity $E[t(y) - \lambda_1'\alpha - \lambda_2'b]^2$ will be referred to as the *mean squared error* (m.s.e.) of the estimator. For any matrix A, A^- will denote any particular generalized inverse of A, i.e., any particular matrix satisfying $AA^-A = A$. Take N to be any matrix whose column space $\mathscr{M}(N)$ is the same as the null space of X'. For any $n \times n$ matrix A, Ω_A will represent the class of those generalized inverses $A^\#$ of A that satisfy the two conditions $\operatorname{rank}(X'A^\#X) = \operatorname{rank}(X)$ and $X'A^\#AN = 0$.

Take $V^\#$ to be any particular member of Ω_V, and let $\hat{\alpha}$ represent any solution to the general normal equations

$$(X'V^\#X)\hat{\alpha} = X'V^\#y \,. \tag{2.1}$$

Suppose that $\lambda_1'\alpha$ is estimable, i.e., that $\lambda_1 = X'\gamma$ for some $n \times 1$ vector γ. A generalized version of the Gauss–Markòv theorem says that Ω_V is nonempty and the linear system (2.1) is consistent, and that $\lambda_1'\hat{\alpha}$ is an essentially-unique best linear unbiased estimator (b.l.u.e.) of $\lambda_1'\alpha$ in the sense that if $c + r'y$ is any other linear unbiased estimator of $\lambda_1'\alpha$, then $\operatorname{Var}(\lambda_1'\hat{\alpha}) \leqq \operatorname{Var}(c + r'y)$ with equality holding if and only if $c + r'y = \lambda_1'\hat{\alpha}$ with probability 1 (see, e.g., Zyskind and Martin (1969), Mitra (1973), and Rao (1973)). Moreover, denoting by $z_{(1)}^*$ and $z_{(2)}^*$ vectors formed from Zyskind and Martin's $z_{(1)}$ and $z_{(2)}$ (see their page 1200) by deleting elements having zero variance; we have that, when y is multivariate normal, $z_{(1)}$, together with $z_{(2)}^{*\prime}z_{(2)}^*$ if σ^2 is unknown, represents a complete sufficient statistic (since otherwise we arrive at a contradiction of the well-known fact that $z_{(1)}^*$, together with $z_{(2)}^{*\prime} z_{(2)}^*$, represents a complete sufficient statistic for the case where the elements of $z_{(1)}^*$ and $z_{(2)}^*$ comprise (all) the observations); so that, when y is normal, $\lambda_1'\hat{\alpha}$ is an essentially-unique best unbiased estimator (b.u.e.) of $\lambda_1'\alpha$.

If α were known, a reasonable estimator for $\lambda_2'\beta$ might be $\lambda_2'\tilde{\beta}$, where $\tilde{\beta} = DZ'V^-(y - X\alpha)$. For α known and V nonsingular, it is well known that $\lambda_2'\tilde{\beta}$ has smaller m.s.e. than any other linear estimator of $\lambda_2'\beta$. When y is normal, $\tilde{\beta}$ is a conditional expectation of b given y, so that, for α known, $\lambda_2'\tilde{\beta}$ then minimizes m.s.e. among all estimators of $\lambda_2'\beta$. These observations suggest that a reasonable estimator for $\lambda_1'\alpha + \lambda_2'\beta$, when α is unknown but $\lambda_1'\alpha$ is estimable, might be $\lambda_1'\hat{\alpha} + \lambda_2'\hat{\beta}$, where

$$\hat{\beta} = DZ'V^-(y - X\hat{\alpha}) .$$

That this estimator can be justified from something more than an intuitive standpoint is demonstrated in the following theorem.

THEOREM 1. *Suppose $\lambda_1'\alpha$ is estimable. The estimator $\lambda_1'\hat{\alpha} + \lambda_2'\hat{\beta}$ is an essentially-unique* b.l.u.e. *of $\lambda_1'\alpha + \lambda_2'\beta$ in the sense that, if $c + r'y$ is any other linear unbiased estimator of $\lambda_1'\alpha + \lambda_2'\beta$, then the* m.s.e. *of $\lambda_1'\hat{\alpha} + \lambda_2'\hat{\beta}$ is less than or equal to the* m.s.e. *of $c + r'y$ with equality holding if and only if $c + r'y = \lambda_1'\hat{\alpha} + \lambda_2'\hat{\beta}$ with probability* 1. *Moreover, when y is normal, $\lambda_1'\hat{\alpha} + \lambda_2'\hat{\beta}$ is an essentially-unique* b.u.e. *of $\lambda_1'\alpha + \lambda_2'\beta$.*

PROOF. Let $u = \lambda_1'\alpha + \lambda_2'b$. For any estimator $t(y)$ possessing a finite second moment,

$$E\{[t(y) - E(u \mid y)][E(u \mid y) - u]\} = E[E\{[t(y) - E(u \mid y)][E(u \mid y) - u] \mid y\}] = 0 ,$$

so that

$$E[t(y) - u]^2 = E[t(y) - E(u \mid y)]^2 + E[E(u \mid y) - u]^2 .$$

Further, $t(y)$ is unbiased for u if and only if $E[t(y) - E(u \mid y)] \equiv 0$. Thus, $t(y)$ is a b.l.u.e. or b.u.e. of u if and only if $t(y) - \lambda_2'DZ'V^-y$ is a b.l.u.e. or b.u.e., respectively, of $(\lambda_1' - \lambda_2'DZ'V^-X)\alpha$, so that it follows from the Gauss–Markov theorem described earlier that $t(y)$ is a b.l.u.e. or, in the case of normality, a b.u.e. of u if and only if $t(y) - \lambda_2'DZ'V^-y = (\lambda_1' - \lambda_2'DZ'V^-X)\hat{\alpha}$ with probability 1. □

When $\lambda_2 = 0$, Theorem 1 reduces of course to the generalized Gauss–Markov theorem. The special case where $\lambda_1 = 0$ and V is nonsingular includes a result due to Henderson (1963).

It is known that $X(X'V^\#X)^-X'V^\#V$ is invariant to the choices for $V^\#$ and $(X'V^\#X)^-$ (a simple proof can be constructed by using Zyskind and Martin's Theorems 1 and 2 together with Lemma 2.2.6(c) from Rao and Mitra's (1971) book). Further, by applying the result $\mathscr{M}(X, V) = \mathscr{M}(X, VN)$ (see, e.g., Rao's (1973) paper) and Rao and Mitra's Lemma 2.2.6(c), it can be shown that, for $y \in \mathscr{M}(X, V)$, $(y - X\hat{\alpha}) \in \mathscr{M}(VN)$. Combining these results with Rao and Mitra's lemma and with the fact that

$$DZ'V^-V = DZ' \tag{2.2}$$

for any choice of V^-, we find that, for $y \in \mathscr{M}(X, V)$, the estimator $\lambda_1'\hat{\alpha} + \lambda_2'\hat{\beta}$ is numerically invariant to the choices for $V^\#$ and V^- and to which solution of (2.1) is used. Note that $\Pr\{y \in \mathscr{M}(X\alpha, V)\} = 1$.

Subsequently, let Λ_1 represent any $p \times t$ matrix such that $\Lambda_1 = X'\Gamma$ for some matrix Γ, so that the elements of $\Lambda_1'\alpha$ are estimable functions of α, and take Λ_2 to be an arbitrary $q \times t$ matrix. The m.s.e.'s of the elements of $\Lambda_1'\hat{\alpha} + \Lambda_2'\hat{\beta}$ lie down the diagonal of the matrix

$$\begin{aligned}\mathrm{Var}\,[\Lambda_1'(\hat{\alpha} - \alpha) + \Lambda_2'(\hat{\beta} - b)] &= \mathrm{Var}\,(\Lambda_1'\hat{\alpha}) + \Lambda_2'[\mathrm{Var}\,(\hat{\beta} - b)]\Lambda_2 \\ &\quad + [\mathrm{Cov}\,(\Lambda_1'\hat{\alpha}, \hat{\beta} - b)]\Lambda_2 + \Lambda_2'[\mathrm{Cov}\,(\Lambda_1'\hat{\alpha}, \hat{\beta} - b)]'\,.\end{aligned}$$

Using (2.2), Zyskind and Martin's Theorems 1 and 2, and Rao and Mitra's Lemma 2.2.6(c), we find

$$\begin{aligned}\sigma^{-2}\,\mathrm{Var}\,(\Lambda_1'\hat{\alpha}) &= \Lambda_1'(X'V^\#X)^-X'V^\#V(V^\#)'X[(X'V^\#X)^-]'\Lambda_1 \\ &= \Lambda_1'(X'V^\#X)^-\Lambda_1\,, \qquad \text{provided } \mathscr{M}(X) \subset \mathscr{M}(V)\,; \\ \sigma^{-2}\,\mathrm{Cov}\,(\Lambda_1'\hat{\alpha}, \hat{\beta} - b) &= -\sigma^{-2}\,\mathrm{Cov}\,(\Lambda_1'\hat{\alpha}, b) = -\Lambda_1'(X'V^\#X)^-X'V^\#ZD\,; \\ \sigma^{-2}\,\mathrm{Var}\,(y - X\hat{\alpha}) &= V - X(X'V^\#X)^-X'V^\#V\,,\end{aligned} \tag{2.3}$$

implying that $X(X'V^\#X)^-XV^\#V$ is symmetric; and

$$\sigma^{-2}\,\mathrm{Var}\,(\hat{\beta} - b) = D - DZ'V^-ZD + DZ'V^-X(X'V^\#X)^-X'V^\#ZD\,.$$

When b is multivariate normal, $E[b \mid (y - X\hat{\alpha})] = E(b \mid \hat{\beta}) = \hat{\beta}$ with probability 1, and $\mathrm{Var}\,[b \mid (y - X\hat{\alpha})] = \mathrm{Var}\,(b \mid \hat{\beta}) = \mathrm{Var}\,(\hat{\beta} - b)$. Thus, the distribution of b, conditional on the vector $(y - X\hat{\alpha})$ of residuals, is the same as the distribution of b, conditional on q linear functions of the residuals which comprise a b.l.u.e. of β.

3. Extended normal equations. Take $b = (b_1', b_2')'$ to be any possible partitioning of b into *uncorrelated* vectors b_1 and b_2 that satisfies the requirement

$$\mathscr{M}(Z_1D_1Z_1') \subset \mathscr{M}(R)\,, \tag{3.1}$$

where $Z = (Z_1, Z_2)$ and $D = \mathrm{diag}\,(D_1, D_2)$ define partitionings of Z and D that correspond in dimension to the partitioning of b and where $R = Z_2D_2Z_2'$ so that $\mathrm{Var}\,(Z_2b_2) = \sigma^2R$.

The representations given in Section 2 for $\Lambda_1'\hat{\alpha} + \Lambda_2'\hat{\beta}$ and for $\mathrm{Var}\,[\Lambda_1'(\hat{\alpha} - \alpha) + \Lambda_2'(\hat{\beta} - b)]$ involve an inverse or generalized inverse for V, a matrix whose dimensions equal the number of data points, which will often be large. Fortunately, in many applications, V has considerable structure which can be exploited to reduce the computations. In particular, there often exists a partitioning of the above type such that R and/or D_1 are diagonal or have other simple forms.

For example, in the case of the ordinary fixed, mixed, or random analysis-of-variance models, we can take b_1 to be a vector whose elements are the random effects associated with the various random factors, and let b_2 represent the vector of errors or residual effects, in which case D_1 is a diagonal matrix whose diagonal elements are ratios of variance components and $R = Z_2 = D_2 = I$ (refer, e.g., to Searle's (1971) book). In this section, we display representations that clearly indicate how such structure can be exploited efficiently for purposes of computing $\Lambda_1'\hat{\alpha} + \Lambda_2'\hat{\beta}$ and Var $[\Lambda_1'(\hat{\alpha} - \alpha) + \Lambda_2'(\hat{\beta} - b)]$.

In what follows, take Q to be any particular matrix satisfying $D_1 = QQ'$ (such a Q always exists), and let S, T and U represent any particular matrices such that $D_1 = STU$. Put

$$V^+ = R^- - R^- Z_1 ST(T + TUZ_1'R^- Z_1 ST)^- TUZ_1'R^- ,$$

take R^* to be any particular member of Ω_R, and put

$$V^* = R^* - R^* Z_1 ST(T + TUZ_1'R^* Z_1 ST)^- TUZ_1'R^* . \tag{3.2}$$

Theorems 2 and 3, to be stated below, reveal that the computations necessary to evaluate $\Lambda_1'\hat{\alpha} + \Lambda_2'\hat{\beta}$ and Var $[\Lambda_1'(\hat{\alpha} - \alpha) + \Lambda_2'(\hat{\beta} - b)]$ are essentially identical to those required to form the linear system

$$\begin{bmatrix} X'R^*X & X'R^*Z_1ST \\ TUZ_1'R^*X & T + TUZ_1'R^*Z_1ST \end{bmatrix} \begin{bmatrix} \tilde{\alpha} \\ \tilde{\phi} \end{bmatrix} = \begin{bmatrix} X'R^*y \\ TUZ_1'R^*y \end{bmatrix} \tag{3.3}$$

and to solve it for $\tilde{\alpha}$ and $\tilde{\phi}$. This relationship is significant because, when R^* can be formed readily by analytical means and/or when D_1, Z_1 and X have many zero elements, it is rather obvious how to exploit these features in forming and solving (3.3). To establish Theorems 2 and 3, we need several lemmas. These lemmas are of at least some interest in themselves.

LEMMA 1.

$$\text{rank}\,(T + TUZ_1'R^-Z_1ST) = \text{rank}\,(T) = \text{rank}\,[T + TUZ_1'R^*Z_1ST + TUZ_1'R^*X(X'R^*X)^-X'R^*Z_1ST] .$$

PROOF. It suffices to prove the second equality, since the first can be viewed as the special case of the second in which X is null. Since $X'R^*RN = 0$, $R(R^*)'X = XF$ for some matrix F; and, since rank $(X'R^*X) =$ rank (X),

$$X(X'R^*X)^-X'R^*X = X \tag{3.4}$$

(see Rao and Mitra's Lemma 2.2.6(c)). Thus,

$$\{R^* - R^*X(X'R^*X)^-X'R^*\}R\{R^* - R^*X(X'R^*X)^-X'R^*\}'R = \{R^* - R^*X(X'R^*X)^-X'R^*\}R . \tag{3.5}$$

Also, (3.1) and the symmetry of D imply that

$$Z_1D_1Z_1' = RM = M'R \tag{3.6}$$

for some matrix M. Using (3.5) and (3.6) together with the well-known result that $\det(I + AB) = \det(I + BA)$ for "any" matrices A and B and the fact that $\det(A) > 0$ for any positive definite matrix A,

$$\begin{aligned}\det[I + UZ_1'\{R^* - R^*X(X'R^*X)^-X'R^*\}Z_1ST] &= \det[I + Q'Z_1'\{R^* - R^*X(X'R^*X)^-X'R^*\} \\ &\quad \times R\{R^* - R^*X(X'R^*X)^-X'R^*\}'Z_1Q] > 0\,,\end{aligned}$$

and the lemma follows. □

LEMMA 2.

$$\begin{aligned}T &= T(T + TUZ'R^-Z_1ST)^-(T + TUZ_1'R^-Z_1ST) \\ &= (T + TUZ_1'R^-Z_1ST)(T + TUZ_1'R^-Z_1ST)^-T\,.\end{aligned}$$

Observing that $T + TUZ_1'R^-Z_1ST = T(T^- + UZ_1'R^-Z_1S)T$, Lemma 2 follows immediately from Lemma 1 and Rao and Mitra's Lemma 2.2.6(f).

LEMMA 3. *The matrix V^+ is a generalized inverse of V;*

$$D_1Z_1'V^+ = ST(T + TUZ_1'R^-Z_1ST)^-TUZ_1'R^-\,; \tag{3.7}$$

$$V^+Z_1D_1 = R^-Z_1ST(T + TUZ_1'R^-Z_1ST)^-TU\,; \tag{3.8}$$

and

$$D_1 - D_1Z_1'V^+Z_1D_1 = ST(T + TUZ_1'R^-Z_1ST)^-TU\,. \tag{3.9}$$

PROOF. Using Lemma 2 together with (3.6),

$$\begin{aligned}V^+V &= R^-R + R^-Z_1ST(T + TUZ_1'R^-Z_1ST)^-TUZ_1'(I - R^-R) \\ &= R^-R + R^-[I - Z_1ST(T + TUZ_1'R^-Z_1ST)^-TUZ_1'R^-] \\ &\quad \times Z_1D_1Z_1'(I - R^-R) \\ &= R^-R\,,\end{aligned} \tag{3.10}$$

so that, again using (3.6), $VV^+V = V$. The result (3.7) follows upon observing that

$$\begin{aligned}TUZ_1'V^+ &= [(T + TUZ_1'R^-Z_1ST) - TUZ_1'R^-Z_1ST] \\ &\quad \times (T + TUZ_1'R^-Z_1ST)^-TUZ_1'R^- \\ &= T(T + TUZ_1'R^-Z_1ST)^-TUZ_1'R^-\,.\end{aligned}$$

Similarly, $V^+Z_1ST = R^-Z_1ST(T + TUZ_1'R^-Z_1ST)^-T$, establishing (3.8). Upon applying Lemma 2, (3.9) follows immediately from (3.7). □

LEMMA 4. *Any partitioned matrix*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{bmatrix}$$

for which $\mathscr{M}(A_{12}) \subset \mathscr{M}(A_{11})$ and $\mathscr{M}(A_{21}') \subset \mathscr{M}(A_{11}')$ has as a generalized inverse

$$\begin{bmatrix} A_{11}^- & 0 \\ 0 & 0\end{bmatrix} + \begin{bmatrix} -A_{11}^-A_{12} \\ I\end{bmatrix} W^-[-A_{21}A_{11}^-, I]\,,$$

where $W = A_{22} - A_{21}A_{11}^-A_{12}$. Moreover, $\operatorname{rank}(A) = \operatorname{rank}(A_{11}) + \operatorname{rank}(W)$.

Lemma 4 represents an extended version of results due to Rohde (1965), and its proof completely parallels Rohde's proofs. There is of course a result analogous to Lemma 4 for the case where $\mathscr{M}(A_{21}) \subset \mathscr{M}(A_{22})$ and $\mathscr{M}(A'_{12}) \subset \mathscr{M}(A'_{22})$. Using (3.4) and the similar result $(X'R^*X)(X'R^*X)^-X' = X'$, it is easy to verify that the conditions of Lemma 4 are met by the partitioned coefficient matrix (subsequently denoted by C) of the linear system (3.3). It follows from Lemma 2 that C also meets the conditions of the analogue to Lemma 4. Applying the analogue, we find that a generalized inverse for C is

$$G^* = \begin{bmatrix} G_{11}^* & G_{12}^* \\ G_{21}^* & G_{22}^* \end{bmatrix}$$

where $G_{11} = (X'V^*X)^-$, $G_{12}^* = -(X'V^*X)^-X'R^*Z_1ST(T + TUZ_1'R^*Z_1ST)^-$, $G_{21}^* = -(T + TUZ_1'R^*Z_1ST)^-TUZ_1'R^*X(X'V^*X)^-$, and $G_{22} = (T + TUZ_1'R^*Z_1ST)^- - G_{21}^*X'R^*Z_1ST(T + TUZ_1'R^*Z_1ST)^-$.

LEMMA 5. *The matrix V^* is a member of Ω_V.*

PROOF. That V^* is a generalized inverse of V is a special case of Lemma 3. Using (3.10), $X'V^*VN = X'R^*RN = 0$. Using Lemmas 1 and 4 and the analogue to Lemma 4, rank (T) + rank $(X'V^*X)$ = rank (C) = rank (T) + rank $(X'R^*X)$, so that rank $(X'V^*X)$ = rank $(X'R^*X)$ = rank (X). □

The following theorem relates the solutions of the system (3.3) to each other, to $\hat{\beta}$ and solutions of (2.1), and thus to b.l.u.e.'s of the elements of the vector $\Lambda_1'\alpha + \Lambda_1'\beta$.

THEOREM 2. *The linear system* (3.3) *is consistent. Suppose that $y \in \mathscr{M}(X, V)$ (which is the case with probability one). Then, for any solution $\hat{\alpha}$ to* (2.1), *the system* (3.3) *has a solution whose first component equals $\hat{\alpha}$. Moreover, if α^* and ψ^* are the components of any solution to* (3.3), *α^* is necessarily a solution to* (2.1) *and $ST\psi^* = \hat{\beta}_1$, where $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')'$. Also (even if the condition* (3.1) *is not satisfied),*

$$\hat{\beta}_2 = D_2Z_2'R^-(y - X\hat{\alpha} - Z_1\hat{\beta}_1)\,.$$

PROOF. Using Lemma 5 and Rao and Mitra's Lemma 2.2.6(c), we find that the system (3.3) has exactly the same solution space as the linear system consisting of the two equations

$$(X'V^*X)\tilde{\alpha} = X'V^*y \tag{3.11}$$

and

$$(T + TUZ_1'R^*Z_1ST)\tilde{\psi} = TUZ_1'R^*(y - X\tilde{\alpha})\,. \tag{3.12}$$

For any $p \times 1$ vector $\tilde{\alpha}$, we can use Lemma 2 to verify that the equation (3.12) is satisfied by

$$\tilde{\psi} = (T + TUZ_1'R^*Z_1ST)^-TUZ_1'R^*(y - X\tilde{\alpha})\,, \tag{3.13}$$

so that the consistency of (3.3) follows from the known consistency of (3.11), and, recalling the invariance properties noted in Section 2, any solution $\hat{\alpha}$ to

(2.1) constitutes the first component of some solution to (3.3) provided $y \in \mathscr{M}(X, V)$. Now suppose that $\tilde{\alpha}$ and $\tilde{\phi}$ are the components of any solution to (3.3) and that $y \in \mathscr{M}(X, V)$. That $\tilde{\alpha}$ is a solution to (2.1) follows from the previously noted invariance properties and the equivalence of the system (3.3) and the system consisting of (3.11) and (3.12). The equivalence of these two linear systems also implies that $\tilde{\phi}$ has the representation (3.13) for some generalized inverse of $(T + TUZ_1'R^*Z_1ST)$, so that, using (3.7) and the invariance properties, $ST\tilde{\phi} = D_1Z_1'V^-(y - X\hat{\alpha}) = \hat{\beta}_1$. Finally, for $y \in \mathscr{M}(X, V)$,

$$\hat{\beta}_2 = D_2Z_2'R^-RV^-(y - X\hat{\alpha}) = D_2Z_2'R^-(y - X\hat{\alpha} - Z_1\hat{\beta}_1)\,. \qquad \square$$

The upshot of Theorem 2 is that the linear system (3.3) can play a role in the extended version, given by Theorem 1, of the generalized Gauss–Markov theorem that is analogous to the role played by the normal equations (2.1) in the original version.

In conjunction with the fixed-effects Gauss–Markov theorem, we know that the covariance matrix of the b.l.u.e.'s of any estimable functions of α can be expressed, through the formula (2.3), in terms of any conditional inverse of the coefficient matrix of the normal equations (2.1). The extension of this result is given by the following theorem.

THEOREM 3. *For any generalized inverse*

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \quad \textit{of} \quad C\,,$$

(3.14)
$$\begin{aligned}\sigma^{-2}\,\mathrm{Var}\,(\Lambda_1'\hat{\alpha}) &= \Lambda_1'G_{11}X'V^*V(V^*)'XG_{11}'\Lambda_1 \\ &= \Lambda_1'G_{11}[I - X'(I - R^*R)(R^*)'XG_{11}']\Lambda_1 \\ &= \Lambda_1'G_{11}\Lambda_1\,, \qquad \textit{provided } \mathscr{M}(X) \subset \mathscr{M}(V)\end{aligned}$$

(3.15)
$$\sigma^{-2}\,\mathrm{Cov}\,(\Lambda_1'\hat{\alpha}, \hat{\beta}_1 - b_1) = \Lambda_1'G_{12}TU\,,$$

(3.16)
$$\begin{aligned}\sigma^{-2}\,\mathrm{Cov}\,(\hat{\beta}_1 - b_1, \Lambda_1'\hat{\alpha}) &= -D_1Z_1'V^*XG_{11}X'V^*V(V^*)'XG_{11}'\Lambda_1 \\ &= STG_{21}[I - X'(I - R^*R)(R^*)'XG_{11}']\Lambda_1 \\ &= STG_{21}\Lambda_1\,, \qquad \textit{provided } \mathscr{M}(X) \subset \mathscr{M}(V)\,,\end{aligned}$$

(3.17)
$$\sigma^{-2}\,\mathrm{Var}\,(\hat{\beta}_1 - b_1) = STG_{22}TU\,.$$

Moreover,

(3.18)
$$\sigma^{-2}\,\mathrm{Cov}\,(\Lambda_1'\hat{\alpha}, \hat{\beta}_2 - b_2) = -\Lambda_1'(G_{11}X' + G_{12}TUZ_1')R^*Z_2D_2\,,$$

(3.19)
$$\sigma^{-2}\,\mathrm{Cov}\,(\hat{\beta}_2 - b_2, \hat{\beta}_1 - b_1) = -D_2Z_2'R^*(XG_{12} + Z_1STG_{22})TU\,,$$

(3.20)
$$\begin{aligned}&\sigma^{-2}\,\mathrm{Var}\,(\hat{\beta}_2 - b_2) \\ &\quad = D_2 - D_2Z_2'R^*Z_2D_2 + D_2Z_2'R^*(XG_{11}X' + XG_{12}TUZ_1' \\ &\qquad + Z_1STG_{21}X' + Z_1STG_{22}TUZ_1')R^*Z_2D_2\,.\end{aligned}$$

PROOF. The results will first be established for the case $G = G^*$. Using (3.10), Lemma 2, and (3.6), we find

$$X'V^{+}V(V^{+})'X = X'V^{+}X - X'R^{-}(I - R^{-}R)'X\,. \tag{3.21}$$

The result (3.14) can be verified readily for $G = G^*$ by applying Lemma 5, the invariance properties noted earlier, (3.21), and Rao and Mitra's Lemma 2.2.6(c). The proof of (3.16) for $G = G^*$ is similar, but makes use of Lemma 3 and the relationship

$$X'V^{\sharp}ZD = X'V^{\sharp}V(V^{\sharp})'X[(X'V^{\sharp}X)^{-}]'X'(V^{-})'ZD\,, \tag{3.22}$$

which follows from (2.2) and the symmetry of $X(X'V^{\sharp}X)^{-}X'V^{\sharp}V$. The verification of (3.15) and (3.17)—(3.20) for $G = G^*$ is made trivial by the availability of Lemma 3.

From a result due to Urquhart (1969), we have that any conditional inverse of C has the representation $G^* + H - G^*CHCG^*$ for some suitably chosen matrix H. Thus, to complete the proof of the theorem, it suffices to show that

$$\begin{bmatrix} X & 0 \\ 0 & T \end{bmatrix} G^*C = \begin{bmatrix} X & 0 \\ 0 & T \end{bmatrix} \tag{3.23}$$

and that

$$CG^*\begin{bmatrix} X' & 0 \\ 0 & T \end{bmatrix} = \begin{bmatrix} X' & 0 \\ 0 & T \end{bmatrix}. \tag{3.24}$$

Using Lemmas 2 and 5 together with Rao and Mitra's Lemma 2.2.6(c), we have that

$$G^*C = \begin{bmatrix} (X'V^*X)^{-}(X'V^*X) & 0 \\ 0 & (T+TUZ_1'R^*Z_1ST)^{-}(T+TUZ_1'R^*Z_1ST) \end{bmatrix}, \tag{3.25}$$

leading to (3.23). The equality (3.24) can be verified in similar fashion. □

In conjunction with the linear system (3.3), we find, using (3.25) along with Lemmas 1 and 5, that

$$\begin{aligned} \operatorname{rank}(C) &= \operatorname{rank}(G^*C) \\ &= \operatorname{rank}[(X'V^*X)^{-}(X'V^*X)] \\ &\qquad + \operatorname{rank}[(T + TUZ_1'R^*Z_1ST)^{-}(T + TUZ_1'R^*Z_1ST)] \\ &= \operatorname{rank}(X'V^*X) + \operatorname{rank}(T + TUZ_1'R^*Z_1ST) \\ &= \operatorname{rank}(X) + \operatorname{rank}(T)\,. \end{aligned}$$

Thus, if $\operatorname{rank}(X) = p$ and if T is chosen to be a nonsingular matrix, then the coefficient matrix C of the linear system (3.3) will be nonsingular even if R and D_1 are not. If $\operatorname{rank}(X) < p$, we could of course consider making a full-rank reparameterization of the fixed effects part of the model in order to achieve a nonsingular coefficient matrix.

Theorems 2 and 3 are generalizations of results due to Henderson (1963). The latter results apply when R and D_1 are both nonsingular, and they are included in the special cases of Theorems 2 and 3 obtained by putting $T = D_1^{-1}$ and $S = U = D_1$. There will typically be values of θ for which D_1 is singular and others for which D_1 is 'ill-conditioned', so that these particular choices for S, T, and U do not always apply and sometimes cause the linear system (3.3) to be unstable numerically. Difficulties of this kind were encountered, for example, by Hemmerle and Hartley (1973).

Some obvious possible choices for T, S, and U that apply even when D_1 is singular are: $T = D_1^-$ and $S = U = D_1$; $T = D_1$ and $S = U = I$; $T = S = I$ and $U = D_1$; $T = U = I$ and $S = D_1$; or $T = I$, $S = Q$, and $U = Q'$. In any given application, S, T, and U should be chosen so that the linear system (3.3) is well-conditioned and, at the same time, easy to form and solve. The results of Zyskind and Martin (1969) and Mitra (1973) are applicable to the computation of R^*.

It is now clear that, for purposes of computing $\Lambda_1'\hat{\alpha} + \Lambda_2'\hat{\beta}$ and Var $[\Lambda_1'(\hat{\alpha} - \alpha) + \Lambda_2'(\hat{\beta} - b)]$, we can readily exploit structure associated with the partitioning $b = (b_1', b_2')'$ by working with the linear system (3.3) rather than with the representations given in Section 2. An alternative approach would be to base the computations on the original representations but to use expressions like (3.2) and (3.7)—(3.9) to advantage whenever possible. It is not hard to see that the latter approach, if properly implemented, is essentially equivalent to the approach based on (3.3).

4. An application of the results in a certain Bayesian setting. Suppose that the statistical model is given by (1.1), but that, instead of all of the elements of b being random effects or errors, some of them, like the elements of α, represent fixed parameters. Put $\alpha^* = \Lambda'\alpha$ where, letting $p^* = \text{rank}(X)$, Λ is any $p \times p^*$ matrix of rank p^* such that $\Lambda = X'\Gamma$ for some Γ, and define X^* to be a $n \times p^*$ matrix of constants such that $X^*\alpha^* \equiv X\alpha$. Conditional on σ^2 and α^*, take the prior distribution of b to be multivariate normal with $E(b) = \mu$ and $\text{Var}(b) = \sigma^2 D$, where μ and D are known. Subsequently, it is supposed that $\mu = 0$ (which can be done without loss of generality); and E^*, Var*, and Cov* are used to indicate expectations, variances, and covariances that are defined with respect to the joint prior distribution of α^*, b and σ^2.

We have

$$E^*(b \mid \alpha^*, y, \sigma^2) = DZ'V^-(y - X^*\alpha^*)\,,$$
$$\text{Var}^*(b \mid \alpha^*, y, \sigma^2) = \sigma^2(D - DZ'V^-ZD)\,,$$

so that

$$E^*(b \mid y, \sigma^2) = DZ'V^-[y - X^*E^*(\alpha^* \mid y, \sigma^2)]\,, \tag{4.1}$$

$$\begin{aligned}\text{Var}^*(b \mid y, \sigma^2) &= E^*[\text{Var}^*(b \mid \alpha^*, y, \sigma^2) \mid y, \sigma^2] \\ &\quad + \text{Var}^*[E^*(b \mid \alpha^*, y, \sigma^2) \mid y, \sigma^2] \\ &= \sigma^2(D - DZ'V^-ZD) \\ &\quad + DZ'V^-X^*[\text{Var}^*(\alpha^* \mid y, \sigma^2)]X^{*\prime}V^-ZD\,,\end{aligned} \tag{4.2}$$

$$\begin{aligned}\text{Cov}^*(b, \alpha^* \mid y, \sigma^2) &= \text{Cov}^*[E^*(b \mid \alpha^*, y, \sigma^2), \alpha^* \mid y, \sigma^2] \\ &= -DZ'V^-X^*[\text{Var}^*(\alpha^* \mid y, \sigma^2)]\,,\end{aligned} \tag{4.3}$$

provided that the prior distribution of α^* and σ^2 is such that the components of α^* and b posses finite second moments.

In circumstances where, conditional on σ^2, there is only vague prior information on the components of α^*, it is to be expected that the conditional posterior mean vector and the conditional posterior covariance matrix of α^*, given σ^2, will be closely approximated by

$$\Lambda'(X'V^{\ddagger}X)^{-}X'V^{\ddagger}y \tag{4.4}$$

and

$$\sigma^2\Lambda'(X'V^{\ddagger}X)^{-}X'V^{\ddagger}V(V^{\ddagger})'X[(X'V^{\ddagger}X)^{-}]'\Lambda\ , \tag{4.5}$$

respectively (see, e.g., Section 11.10 in DeGroot's (1970) book). Moreover, formal substitution of (4.4) and (4.5) in the right hand sides of (4.1)—(4.3) yields corresponding approximations for $E^*(b \mid y, \sigma^2)$, Var* $(b \mid y, \sigma^2)$, and Cov* $(b, \alpha^* \mid y, \sigma^2)$. Recalling (3.22), it is clear that the manipulations required to compute these approximations are exactly the same as those associated with the best linear unbiased estimation of α^* and β, so that the results of Section 3 can be used to facilitate these manipulations.

There are at least three settings where we might want to use a prior distribution that reflects only vague prior information on certain of the parameters (corresponding to the components of α^*) but which indicates significant prior information on others (those that are components of b):

1. There truly may be little prior information on some of the parameters. Lindley and Smith (1972) indicated in effect that the formulation of a suitable prior is often facilitated by introducing hyperparameters, for which a vague prior may be satisfactory, and by expressing some or all of the original parameters as deviations from these hyperparameters. The model, after its reparameterization in this way, will contain parameters for which a vague prior is appropriate even though the original model does not.

2. Only a relatively few of the parameters (corresponding to components of b) may be of interest, with the rest (the components of α^*) being regarded as nuisance parameters. There may be important prior information on the parameters of interest which we wish to incorporate. There may also be significant prior information on α^*, but it may be hard to express in the form of a probability distribution, so that the added precision resulting from its use may not be worth the trouble.

3. Suppose that an important decision is to be made on the basis of the posterior distribution of certain of the parameters (which we take to be the components of α^*). Suppose also that the decision involves competing interests, and consequently agreement cannot be reached on an appropriate prior distribution for α^*. Then, it might be agreed to make the decision on the basis of a posterior distribution appropriate for someone who is ignorant of any prior information on α^*, yet at the same time we might want to take advantage of prior information on other of the parameters (those that are components of b).

Acknowledgments. The author is indebted to an Associate Editor and two referees, whose suggestions and constructive criticisms prompted considerable generalization and consolidation of the original results.

REFERENCES

[1] DeGroot, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

[2] Hemmerle, W. J. and Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the *W* transformation. *Technometrics* **15** 819–831.

[3] Henderson, C. R. (1963). Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding,* National Academy of Sciences-National Research Council Publication No. 982, 141–163.

[4] Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *J. Roy. Statist. Soc. Ser. B* **34** 1–18.

[5] Mitra, S. K. (1973). Unified least squares approach to linear estimation in a general Gauss–Markov model. *SIAM J. Appl. Math.* **25** 671–680.

[6] Rao, C. R. (1973). Representations of best linear unbiased estimators in the Gauss–Markoff model with a singular dispersion matrix. *J. Multivariate Anal.* **3** 276–292.

[7] Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications.* Wiley, New York.

[8] Rohde, C. A. (1965). Generalized inverses of partitioned matrices. *SIAM J. Appl. Math.* **13** 1033–1035.

[9] Searle, S. R. (1971). *Linear Models.* Wiley, New York.

[10] Searle, S. R. (1974). Prediction, mixed models, and variance components. *Reliability and Biometry* (F. Proschan and R. J. Serfling, eds.) 229–266. SIAM, Philadelphia.

[11] Urquhart, N. S. (1969). The nature of the lack of uniqueness of generalized inverse matrices. *SIAM Rev.* **11** 268–271.

[12] Zyskind, G. and Martin, F. B. (1969). On best linear estimation and a general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM J. Appl. Math.* **17** 1190–1202.

Department of Statistics
Iowa State University
Ames, Iowa 50010